Two-Sample Inference Procedures for Population Central Values

Josh Jarvey

05/26/2020

## Exercise 1

An experiment was conducted to evaluate the effectiveness of a drug treatment for tapeworm in the stomachs of sheep. A random sample of 24 worm-infected lambs of approximately the same age was randomly divided into two groups. Twelve of the lambs were injected with the drug and the remaining twelve were left untreated. After a 6-month period the lambs were slaughtered and the worm counts recorded.

### Part 1a

Load the dataset WormSheep from the DS705data package. Note that the package is loaded above at line 18. You just need the data() command. To see all the data sets in the package, type data(package=‘DS705data’).

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1a -|-|-|-|-|-|-|-|-|-|-|-

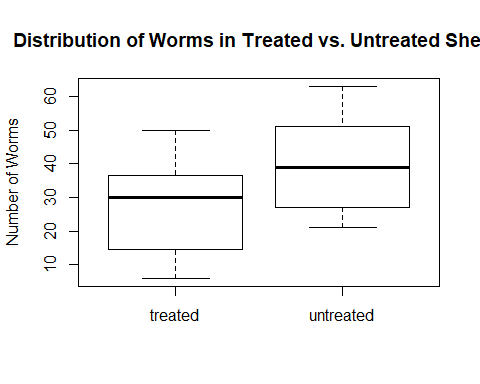
#load in the WormSheep dataset from the DS705data package  
data("WormSheep")

### Part 1b

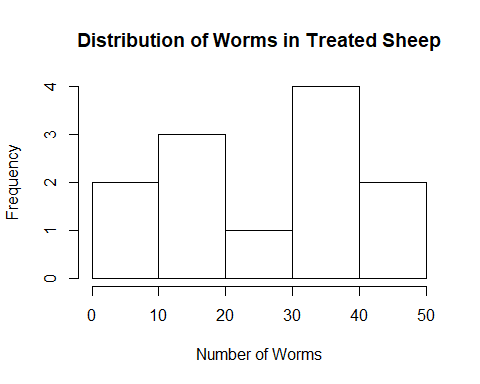
Create boxplot, histograms, and normal probability plots for each group (treated vs. untreated). Be sure that each plot is labeled appropriately.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1b -|-|-|-|-|-|-|-|-|-|-|-

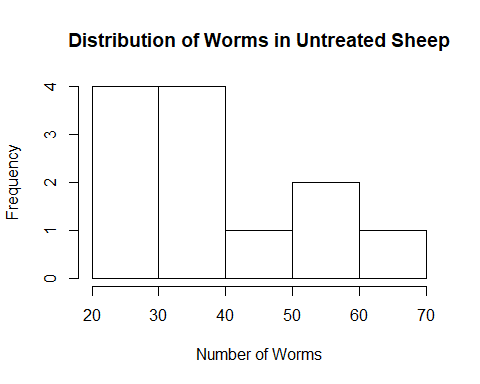
#create a boxplot of worms by treatment type.  
boxplot(worms~treated, data = WormSheep, main = "Distribution of Worms in Treated vs. Untreated Sheep", xlab = '', ylab = "Number of Worms")



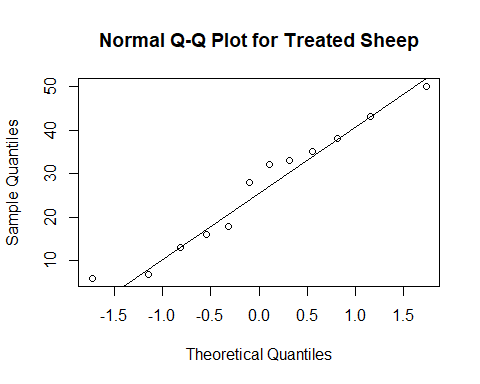
#pull the appropiate data into their own vectors - treated vs. untreated  
treated = WormSheep[WormSheep$treated == "treated",1]  
untreated = WormSheep[WormSheep$treated == "untreated",1]  
  
 #create histograms of the treated and untreated sheep  
hist(treated, main = "Distribution of Worms in Treated Sheep", xlab = "Number of Worms")



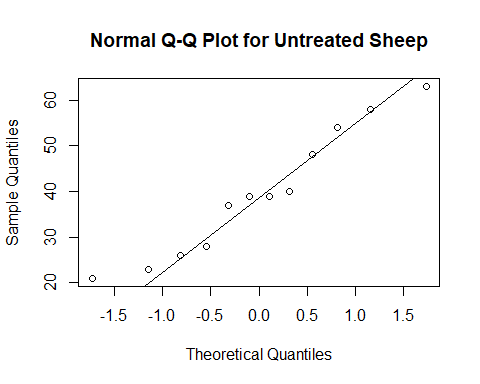
hist(untreated, main = "Distribution of Worms in Untreated Sheep", xlab = "Number of Worms")



#create QQ plots to assess normality in the samples  
qqnorm(treated, main = "Normal Q-Q Plot for Treated Sheep")  
qqline(treated)



qqnorm(untreated, main = "Normal Q-Q Plot for Untreated Sheep")  
qqline(untreated)



### Part 1c

Do the boxplots show any outliers?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1c -|-|-|-|-|-|-|-|-|-|-|-

In reviewing the boxplots, there do not appear to be any outliers in either samples.

### Part 1d

Describe the shapes of the histograms for the sample data for each sample.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1d -|-|-|-|-|-|-|-|-|-|-|-

The treated sheep sample appears to be bivariate, with two bumps on either end of the sample There does not appear to be any skew to the data.

The untreated sheep sample appears to have a heavy number of observations on the left hand side of the sample, and the data appears to be skewed to the right.

### Part 1e

Do either of the normal probability plots show evidence of non-normality? Explain.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1d -|-|-|-|-|-|-|-|-|-|-|-

The observations in both samples generally follow the theoritical normal distribution. There’s slight indication of right-skew in the untreated group (there are a few observations hovering above the line in the lower tail). But it’s fair to say they do not show evidence of non-normality.

### Part 1f

Conduct an appropriate test to determine if the worm counts in each population can be considered as normally distributed. Provide the p-value and the conclusion of the test at a 5% level of significance.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1f -|-|-|-|-|-|-|-|-|-|-|-

#using the shapiro test to test the normality of the sample for treated sheep.  
shapiro.test(treated)

##   
## Shapiro-Wilk normality test  
##   
## data: treated  
## W = 0.95092, p-value = 0.6504

#using the shapiro test to test the normality of the sample for untreated sheep.  
shapiro.test(untreated)

##   
## Shapiro-Wilk normality test  
##   
## data: untreated  
## W = 0.94382, p-value = 0.5491

At a 95% significance level, there is not enough evidence to reject that the sample of “treated” sheep from the WormSheep dataset is normally distributed (p-value=0.6504).

## At a 95% significance level, there is not enough evidence to reject that the sample of “untreated” sheep from the WormSheep dataset is normally distributed (p-value=0.5491).

### Part 1g

Conduct the test of your choice to determine if the population mean worm count for all sheep treated with the drug is less than the mean worm count for the population of untreated sheep. Let .

#### Step 1

State the null and alternative hypotheses for the test. If you use symbols, please define them.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g.step1 -|-|-|-|-|-|-|-|-|-|-|-

$$ \mu\_\mbox{1} \ = Population\ mean\ of\ treated\ sheep.\\
\mu\_\mbox{2} \ = Population\ mean\ of\ untreated\ sheep.\\
H\_0: \mu\_\mbox{1} \ >= \mu\_\mbox{2}\\
H\_a: \mu\_\mbox{1} \ < \mu\_\mbox{2}$$

#### Step 2

Use R to generate the output for the test you selected.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g.step2 -|-|-|-|-|-|-|-|-|-|-|-

#conducting t-test.  
t.test(treated,untreated,alternative = "less")

##   
## Welch Two Sample t-test  
##   
## data: treated and untreated  
## t = -2.2709, df = 21.972, p-value = 0.01665  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -3.189613  
## sample estimates:  
## mean of x mean of y   
## 26.58333 39.66667

#### Step 3

State both a statistical conclusion at and interpret it in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g.step3 -|-|-|-|-|-|-|-|-|-|-|-

At a 95% significance level, there is enough evidence to support the claim that sheep treated with the drug have less worms on average than those sheep who are not treated (p-value=0.01665)

### Part 1h

Compute a 95% confidence interval for the difference in population means. Write an interpretation in the context of the problem for this interval.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1h -|-|-|-|-|-|-|-|-|-|-|-

#conducting a t-test for difference in means, testing mu1 - mu2.   
t.test(treated,untreated,alternative = "two.sided")

##   
## Welch Two Sample t-test  
##   
## data: treated and untreated  
## t = -2.2709, df = 21.972, p-value = 0.03331  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -25.032642 -1.134025  
## sample estimates:  
## mean of x mean of y   
## 26.58333 39.66667

With 95% confidence, the population mean worm count in treated sheep is 25.58 to 1.13 less worms than the population mean worm count of sheep without treatment.

### Part 1i

Did you use the Welch t-test or the Wilcoxon rank sum test? Justify your choice, including some discussion of how well the conditions for the hypothesis test and confidence interval procedures were met.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1i -|-|-|-|-|-|-|-|-|-|-|-

I used the Welch t-test for the testing of population means in question 1. While both tests assume a simple random sample, it was stated in the intro that these sheep were randomly selected and randomly divided. However, I chose the Welch test over the Wilcoxon rank sum test because the samples were approximately normally distributed, and the sample size was above 10 in both groups. There were also no outliers in the sample, and although the sample of untreated was mildly skewed - the Welch test will still perform fine.

## Problem 2

Consider two independent, random samples of delay times (in hour) for two airlines over the past 10 years. The task here is to compare the population central values by conducting a hypothesis test and constructing a 95% confidence interval for the difference in average delay time between the airlines for the population of all delays in the past 10 years.

### Part 2a

Load the dataset delays from the DS705data package.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2a -|-|-|-|-|-|-|-|-|-|-|-

#loading the "delays" dataset from the DS705data package.  
data("delays")

### Part 2b

Begin by exploring the sample means, medians, and standard deviations for each airline. That is, find the sample means, medians, and standard deviations.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2b -|-|-|-|-|-|-|-|-|-|-|-

#pulling delay times into their own vectors, by airport  
omega = delays[delays$airline == "Omega",1]  
skybird = delays[delays$airline == "Skybird",1]  
  
 #calculating the mean, median and sd for omega  
summary(omega)[3:4]

## Median Mean   
## 1.0850 1.4084

print(paste("Standard Devidation:",sd(omega)))

## [1] "Standard Devidation: 1.05821250903857"

#calculating the mean, median and sd for skybird  
summary(skybird)[3:4]

## Median Mean   
## 1.260000 1.421333

print(paste("Standard Devidation:",sd(skybird)))

## [1] "Standard Devidation: 0.946192369447144"

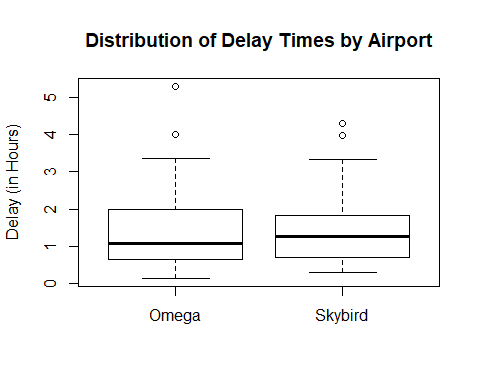
### Part 2b

Produce boxplots, histograms, and normal probability plots for the delay times of both airlines.

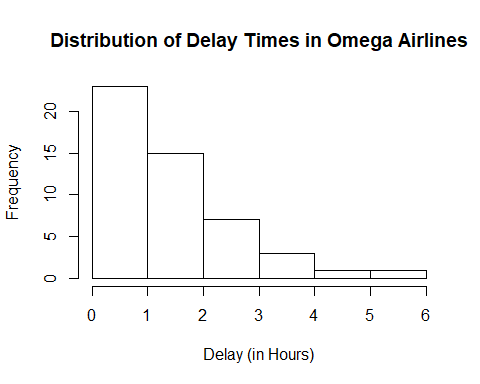
Discuss the aspects of the plots that are relevant to the construction of the best type of test and confidence interval to compare the population central values.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2b -|-|-|-|-|-|-|-|-|-|-|-

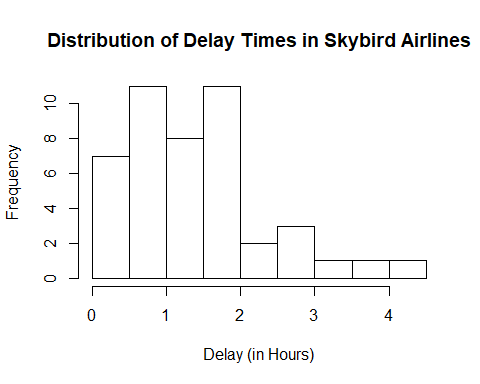
#create a boxplot of delay times by airline.   
boxplot(delay~airline, data = delays, main = "Distribution of Delay Times by Airport", xlab = '', ylab = "Delay (in Hours)")



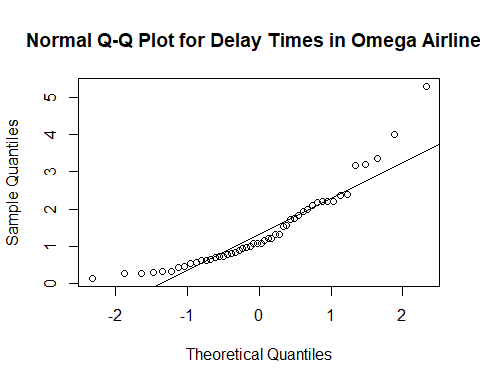
#create histograms of the delay times of omega and skybird  
hist(omega, main = "Distribution of Delay Times in Omega Airlines", xlab = "Delay (in Hours)")



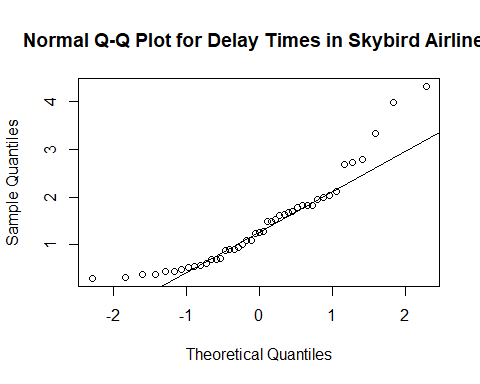
hist(skybird, main = "Distribution of Delay Times in Skybird Airlines", xlab = "Delay (in Hours)")



#create QQ plots to assess normality in the samples  
qqnorm(omega, main = "Normal Q-Q Plot for Delay Times in Omega Airlines")  
qqline(omega)



qqnorm(skybird, main = "Normal Q-Q Plot for Delay Times in Skybird Airlines")  
qqline(skybird)



Because the mean is greater than the median when reviewing the summary statistics of both sample data sets, this indicates a right-skew. This is further corroborated by reviewing the boxplot (shows outliers), histograms, and QQ Plots of both sample sets. Since the data demonstrates these characteristics, the Wilcoxon Rank Sum test will be more appropiate here.

### Part 2c

Conduct hypothesis tests for the normality of the delay times for both Skybird and Omega using a 5% level of significance in each case.

Use R to compute the test statistic and P-value.

State the hypotheses and conclusion of the test. Include the P-value in your conclusion.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2c -|-|-|-|-|-|-|-|-|-|-|-

#perform a shapiro test to assess the normality of the omega sample.  
shapiro.test(omega)

##   
## Shapiro-Wilk normality test  
##   
## data: omega  
## W = 0.87099, p-value = 6.027e-05

#perform a shapiro test to assess the normality of the skybird sample.  
shapiro.test(skybird)

##   
## Shapiro-Wilk normality test  
##   
## data: skybird  
## W = 0.89362, p-value = 0.0006031

$$H\_0: \ The\ sample\ of\ delay\ times\ from\ Omega\ airlines\ was\ drawn\ from\ a\ normally\ distributed\ population.\\
H\_a: \ The\ sample\ of\ delay\ times\ from\ Omega\ airlines\ was\ not\ drawn\ from\ a\ normally\ distributed\ population.$$

At a 95% signficance level, there is enough evidence to conclude that the distribution of delay times for the last 10 year’s worth of flights from Omega Airlines are not normally distributed (p-value = 6.027e-05).

$$H\_0: \ The\ sample\ of\ delay\ times\ from\ Skybird\ airlines\ was\ drawn\ from\ a\ normally\ distributed\ population.\\
H\_a: \ The\ sample\ of\ delay\ times\ from\ Skybird\ airlines\ was\ not\ drawn\ from\ a\ normally\ distributed\ population.$$

At a 95% signficance level, there is enough evidence to conclude that the distribution of delay times for the last 10 year’s worth of flights from Skybird Airlines are not normally distributed (p-value = 6.031e-04).

### Part 2d

Perform the most appropriate test to compare population central values for all Skybird and Omega delay times for the past 10 years. Use a 10% level of significance.

#### Step 1

State the null and alternative hypotheses for the test (in words).

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d.step1 -|-|-|-|-|-|-|-|-|-|-|-

$$H\_0: \ The\ distributions\ of\ delay\ times\ are\ identical\ for\ both\ Omega\ and\ Skybird\ airlines.\\
H\_a: \ The\ distributions\ of\ delay\ times\ are\ shifted\ from\ each\ other\ for\ both\ Omega\ and\ Skybird\ airlines.$$

#### Step 2

Use R to generate the output for the test you selected.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d.step2 -|-|-|-|-|-|-|-|-|-|-|-

#perform a wilcoxon rank sum test to assess if the distributions between omega and skybird are different  
wilcox.test(omega,skybird,conf.level = 0.90)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: omega and skybird  
## W = 1093.5, p-value = 0.8173  
## alternative hypothesis: true location shift is not equal to 0

#### Step 3

State both a statistical conclusion at and interpret it in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d.step3 -|-|-|-|-|-|-|-|-|-|-|-

At a 90% significance level, there is not enough evidence to claim that the distribution of delay times for the last 10 year’s worth of flights between Omega and Skybird airlines are shifted from each other (p-value = 0.8173).

### Part 2e

Construct the most appropriate 95% confidence interval to compare population central values. Write a sentence to interpret the interval in the context of the problem.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2e -|-|-|-|-|-|-|-|-|-|-|-

#perform a wilcoxon rank sum test to calculate the confidence interval at the 95% significance level.   
wilcox.test(omega,skybird,conf.level = 0.95, conf.int = TRUE)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: omega and skybird  
## W = 1093.5, p-value = 0.8173  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## -0.3900065 0.2800410  
## sample estimates:  
## difference in location   
## -0.04997005

With 95% confidence, the shift in distributions of delay times for the last 10 year’s worth of flights from Omega and Skybird airlines is between -0.39 and 0.28 hours.

### Part 2f

Did you use the Welch t-test and interval or the Wilcoxon rank sum test with the corresponding interval? Justify your choice, including some discussion of how well the conditions for the hypothesis test and confidence interval procedures were met.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2f -|-|-|-|-|-|-|-|-|-|-|-

I decided to use the Wilcoxon Rank Sum test in this exercise because the samples were not normally distributed, which is a requirement for the Welch t-test. The distribution of the samples did appear approximately similar in shape, further indicating support for the Wilcoxon test.

### Part 2g

What type of hypothesis testing error could have been made here? Describe it in the context of the problem.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2g -|-|-|-|-|-|-|-|-|-|-|-

Because we did not reject the null, we have the chance of creating a Type II error if the null is actually false in reality. This would mean that the difference in population median delay times for the last 10 year’s worth of flights between Omega and Skybird were actually different (i.e. the confidence interval didnt include 0 within its range).

### Part 2h

Construct a 95% bootstrap confidence interval to compare population medians for the delay times over the past 10 years for the two airlines. Use the Bca method and a bootstrap sample size of 2000. Compare it to the interval produced by the wilcox.test function.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2h -|-|-|-|-|-|-|-|-|-|-|-

#import required packages  
library(boot)

## Warning: package 'boot' was built under R version 3.6.3

#create an auxilliary function to caculate the difference in medians.  
bootMedianDiff <- function(d,i){  
 medians = tapply(d[i,1], d[,2], median)  
 medians[1]-medians[2]  
}  
  
 #complete the bootstrapping to create a sampling distribution of 2000 median differences using the delay times. Group by airline.   
boot.object = boot(delays, bootMedianDiff, R = 2000, strata = delays$airline)  
  
 #create a 95% bca confidence interval using the bootstrapped data.  
boot.ci(boot.object, conf = 0.95, type = 'bca')

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS  
## Based on 2000 bootstrap replicates  
##   
## CALL :   
## boot.ci(boot.out = boot.object, conf = 0.95, type = "bca")  
##   
## Intervals :   
## Level BCa   
## 95% (-0.69, 0.32 )   
## Calculations and Intervals on Original Scale

The 95% confidence interval using the bootstrap medians and BCa method is wider than the Wilcoxon Rank Sums test.